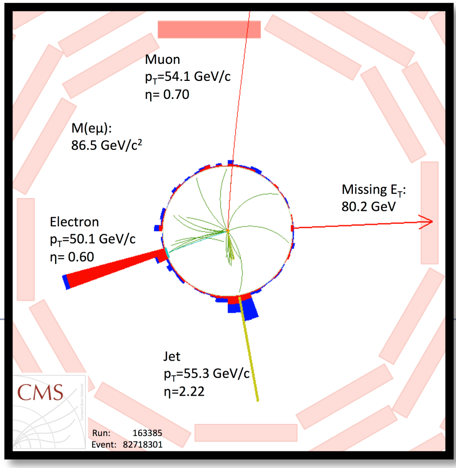
4th generation and single-top:
Workshop on fourth fermion generation and on single-top production
26-28 Mar 2012



tW production at CMS

Rebeca Gonzalez Suarez
on behalf of the CMS collaboration

Intro

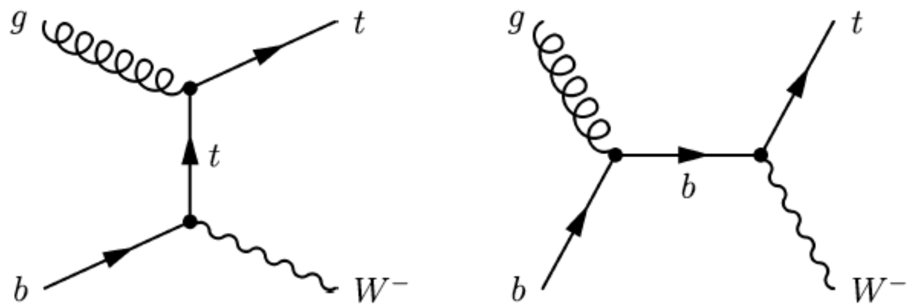
- CMS's first public results for tW with 7 TeV data (September 2011):
  CMS PAS TOP-11-022

- **First** experimental study of the Single Top Associated tW in CMS:
  - **Too difficult at the Tevatron**, with no results from CDF and D0
  - At the LHC, **ATLAS** had already public results

- Very **interesting signature:**
  - important background to Higgs searches ($H \rightarrow WW$, particularly 1jet bin)
- Not yet observed
- After the observation of the process, **a lot of physics to do:**
  - **Wtb interaction vertex** study in a complementary region to the ones of the s- or t-channel production
  - direct measurement of the **$|V_{tb}|$ CKM matrix element** without assumptions about the number of quark generations

tW dilepton topology

- Three final states **(ee/eμ/μμ)** are studied:



Signal signature

2 isolated opposite charged leptons
Missing E_T (2 neutrinos)
1 jet (coming from a b-decay)

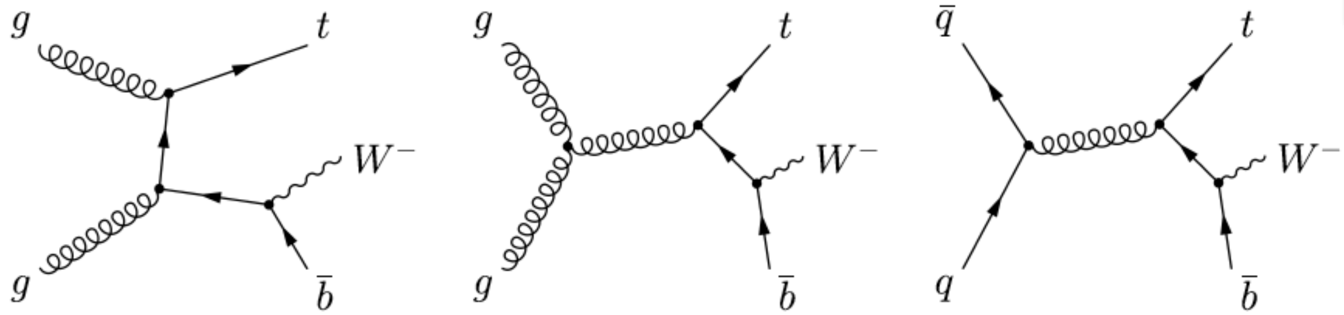
- All the processes able to produce similar signatures will be background sources:
  - **ttbar** (dominant background), **Z+jets** (ee/μμ), EWK di-boson production (**WW, WZ, ZZ**), **W+jets**, other single top processes (**t and s channel**) and **QCD**

Process	Cross-section
tW	15.6±0.4+1.0-1.2 pb (app. NNLO)
tt	163±14 pb (app. NNLO)
Z+jets	3048±132 pb (NNLO)

Diagram Removal and Subtraction

The definition of the tW process in perturbative QCD **mixes with top-pair production (tt) at NLO**, and poses conceptual problems

NLO



(a) gluon-box (b) gluon-fusion (c) $q\bar{q}$

Two possible solutions:

-The "diagram removal" approach (DR): all ambiguous diagrams at NLO are excluded from the definition of signal (**chosen as default in the analysis)**

- The "diagram subtraction" approach (DS): subtracts a gauge-invariant term, cancelling locally the contribution of tt diagrams

The differences of using one or another Monte Carlo are small at the end of the analysis chain and taken as a **systematic uncertainty**

Datasets and Trigger

- $2.1 fb^{-1}$ of data considered, corresponding to 2011 data taken before September
- CMS has 4.98 ± 0.11 fb^{-1} of 7TeV proton-proton collisions available, being currently analyzed

- Dilepton final states studied:
  - **dilepton data streams**, **dilepton HLT paths**:
    - 1e1μ, pt > 17/8 (8/17)
    - 2μ, pt > 17/8 (13/8) (7/7)
    - 2e, pt > 17/8
- The efficiency is extracted from data and emulated in MC. **Scale factors** from data/simulation comparison applied

Scale factors for the trigger (PAS **TOP-11-005**):
0.961±0.01 (μμ) / 0.977±0.025 (ee) / 0.987±0.020 (eμ)

- **Monte Carlo**:
  - Single-top events in all channels simulated with **POWHEG** (full NLO)
  - **MADGRAPH** is used for tt, Z+jets, W+jets, **PYTHIA** for the rest
  - CTEQ 6.6 PDF sets

Physics objects: Leptons

Muons:
- Reconstructed using information from the Tracker and Muon Chambers
- $p_T > 20$ GeV
- $|\eta| < 2.4$
- Standard CMS requirements involving the quality of the reconstructed track, number of hits and impact parameter
- Isolation variable (sum of the transverse momentum of charged and neutral particles divided by the lepton P_T) in $\Delta R \leq 0.3 < 0.15$

Loose Muons:
- $p_T > 10$ GeV, $|\eta| < 2.5$, iso. < 0.2

Electrons:
- Reconstructed from clusters of energy deposits in the ECAL matched to hits in the silicon tracker
- $E_T > 20$ GeV
- $|\eta| < 2.5$
- Cut-based selection cuts using shower shape and track-cluster matching to reject fakes maintaining at least 70% efficiency for electrons from the decay of W (WP70)
- Quality cuts on the impact parameter, distance to the PV, as well as cuts against photon conversions
- Isolation (same as muons) < 0.15

Loose Electrons:
- $E_T > 15$ GeV, $|\eta| < 2.5$, WP95, iso. < 0.2

Data-driven scale factors for the lepton identification and isolation (PAS **TOP-11-005**):
0.947±0.002 μμ /0.968±0.003 eμ /0.990±0.001 ee

Physics objects: Jets and Missing E_T

Jets:
- Anti-kt jets (0.5)
- Corrected P_T > 30 (20 for loose jets) GeV
- $|\eta| < 2.4$
- standard Jet identification criteria applied
- $\Delta R_{jet\text{-}lepton} > 0.3$

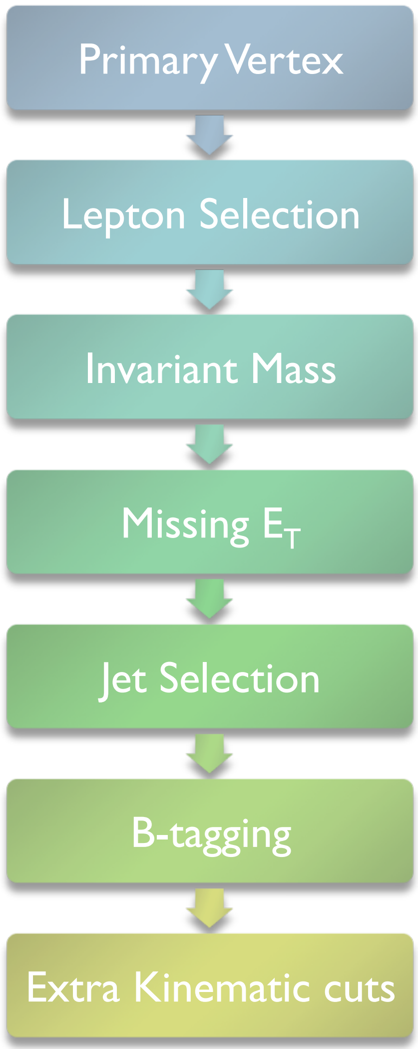
Missing ET:
- Standard Particle Flow Missing E_T

B-tagging:
- Algorithm reconstructs decay vertex and requires a minimum number of associated tracks: **Simple Secondary Vertex**
- The working point is High efficiency Medium (SSVHEM, disc. > 1.74)
- b-tagging efficiency of 62% with a mistag rate of 1.43%

Data-driven scale factor b-tagging (PAS **BTV-11-001**): 0.95±0.01(stat)±0.10(sys)

- Leptons, jets and missing E_T are reconstructed by the **Particle Flow** (PF) algorithm, which performs a **global event reconstruction** providing the full list of particles identified as **electrons, muons, photons, charged and neutral hadrons**

Analysis Flow



Analysis Flow (II)

- **Cleaning:** (data) Events with very high energy noise in the HCAL are rejected; events with 10 tracks or more, should have at least 25% of them passing a tight selection
- **Primary Vertex cut:** At least one primary vertex with 4 tracks or more, with longitudinal (radial) distance of less than 24 (2) cm from the center of the detector
- **Lepton Selection:** exactly two isolated leptons (electrons or muons) with opposite charge, events with other leptons (loose) are vetoed
- **Invariant Mass and MET cuts:**
  - At pre-selection level, all final states: $m_{ll} > 20$ GeV
  - **ee/μμ final states**:
    - Events with invariant mass m_{ll} inside the Z mass window [81, 101] are rejected
    - Missing $E_T > 30$ GeV

Mostly against Drell-Yan background (also other Z processes and QCD)

Analysis Flow (III)

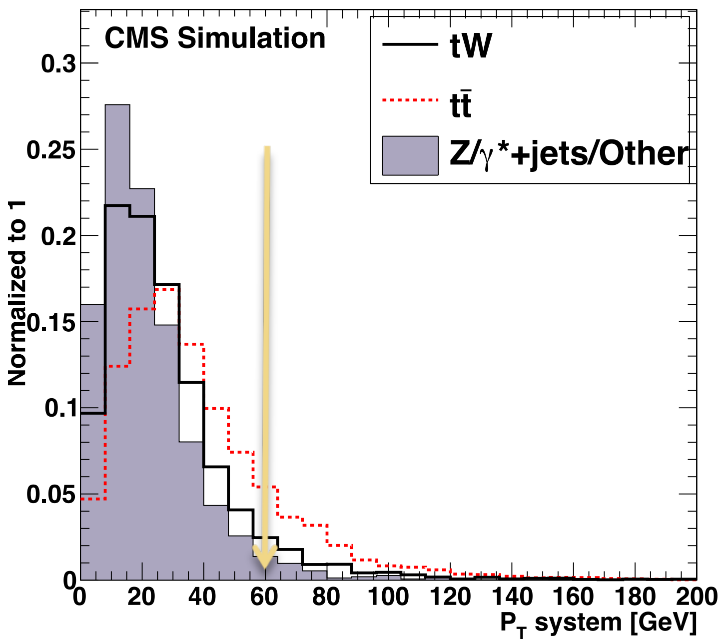
- **Jet selection:** To define the **Signal Region**, events are required to have exactly 1 jet, following the requirements presented previously, with $p_T > 30$ GeV
- **B-tagging:** The jet in the event has to be b-tagged, and no extra b-tagged jets are allowed ($p_T > 20$ GeV)

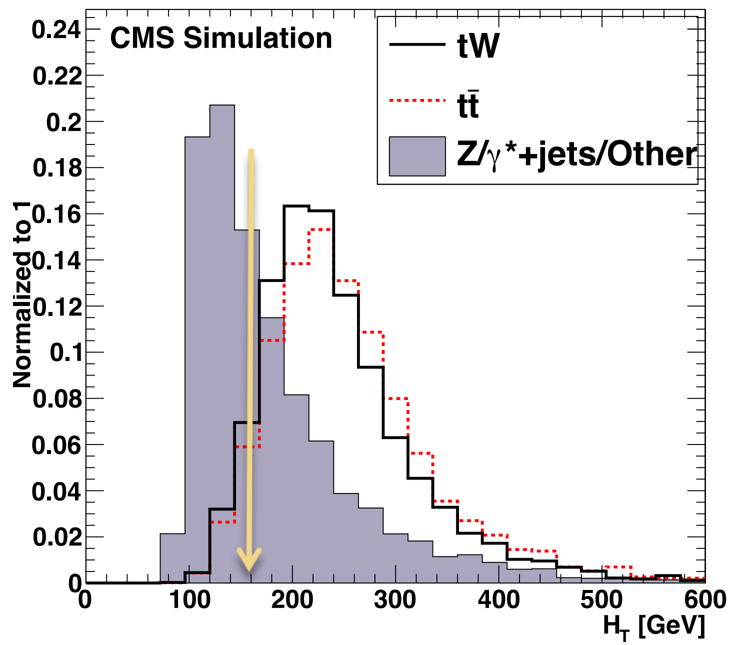
Other Jet multiplicities are also used to constrain tt background

- **Extra kinematical cuts:** chosen maximizing the signal significance after b-tagging (against tt), in the p_T of the system (leptons + jet + MET) and, only in the eμ channel, to compensate for the lack of MET and m_{ll} cuts, in H_T

Variable	cut
p_T of the system	< 60 GeV
Ht	> 160 GeV (eμ only)

p_T of the system and H_T





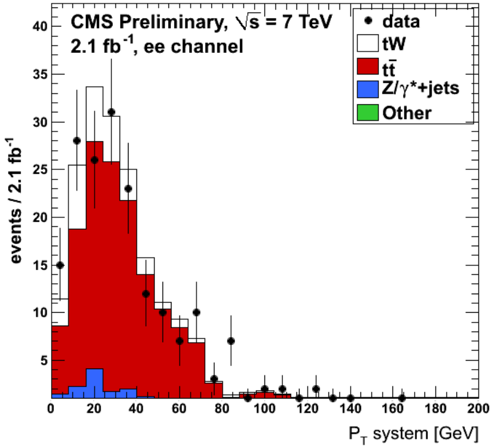
Kinematic distributions normalized to the same area in the eμ final state for simulated tW signal, tt background and the sum of other backgrounds after lepton selection and lepton veto, in events with exactly one jet.

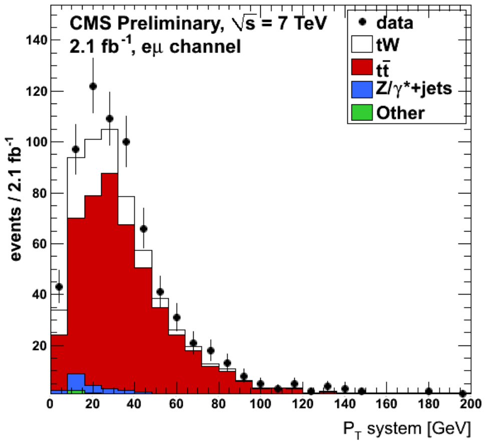
Summary of the selection

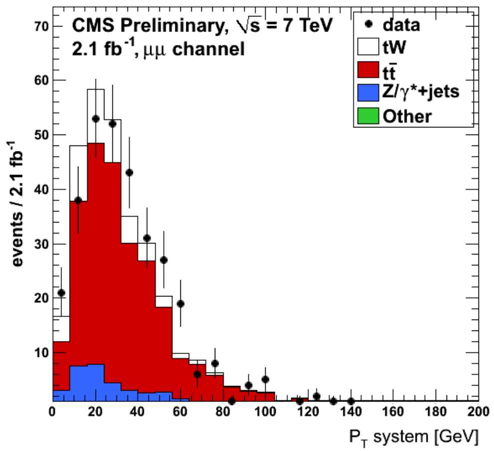
Cut	ee	eμ	μμ
HLT	yes	yes	yes
Primary Vertex (standard cleaning cuts)	yes	yes	yes
Opposite charge 2 leptons selection $p_T > 20$ GeV	yes	yes	yes
Loose lepton veto (no extra loose leptons)	yes	yes	yes
$m_{ll} > 20$ GeV	yes	yes	yes
m_{ll} outside [81,101] GeV	yes	no	yes
MET > 30 GeV	yes	no	yes
1 jet selection ($p_T > 30$ GeV)	yes	yes	yes
B-tagging requirements (SSHVEM > 1.74, veto extra b-tagged $p_T > 20$ GeV)	yes	yes	yes
p_T of the system < 60 GeV	yes	yes	yes
$H_T > 160$ GeV	no	yes	no

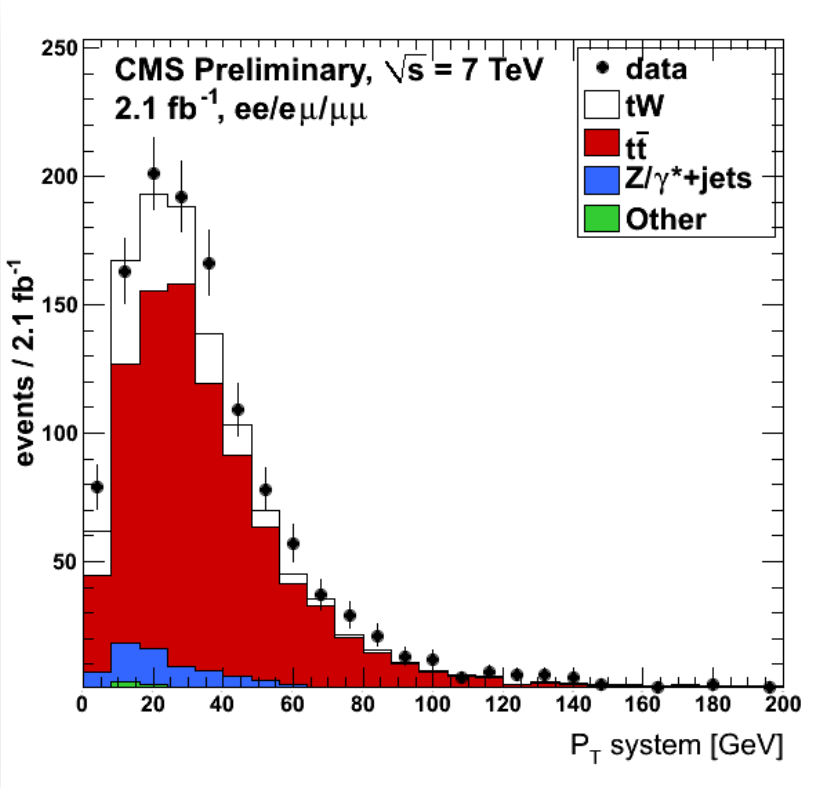
In Data: Event cleaning
In Monte Carlo: Pile Up re-weighting to match the data
HLT and lepton ID and isolation scale factors applied per event
B-tagging scale factor applied per jet

Final MC/data distributions: P_T system



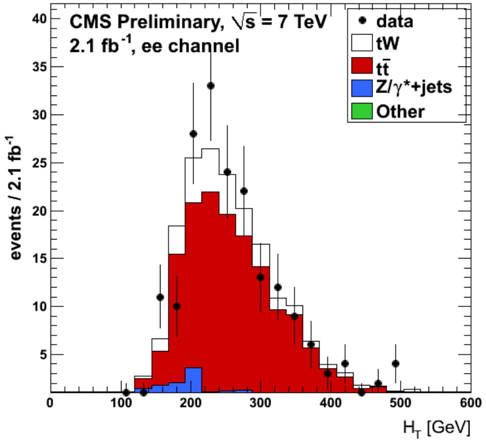


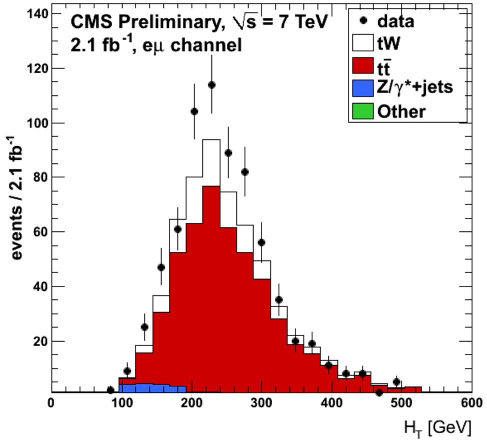


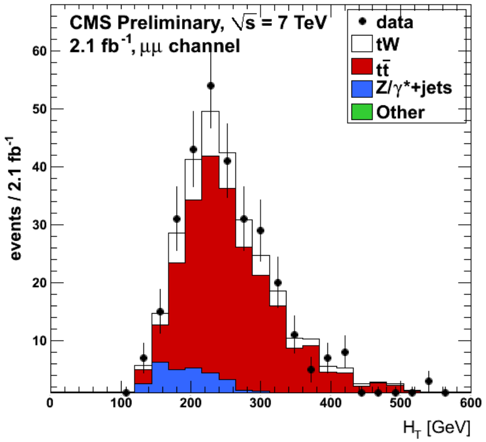


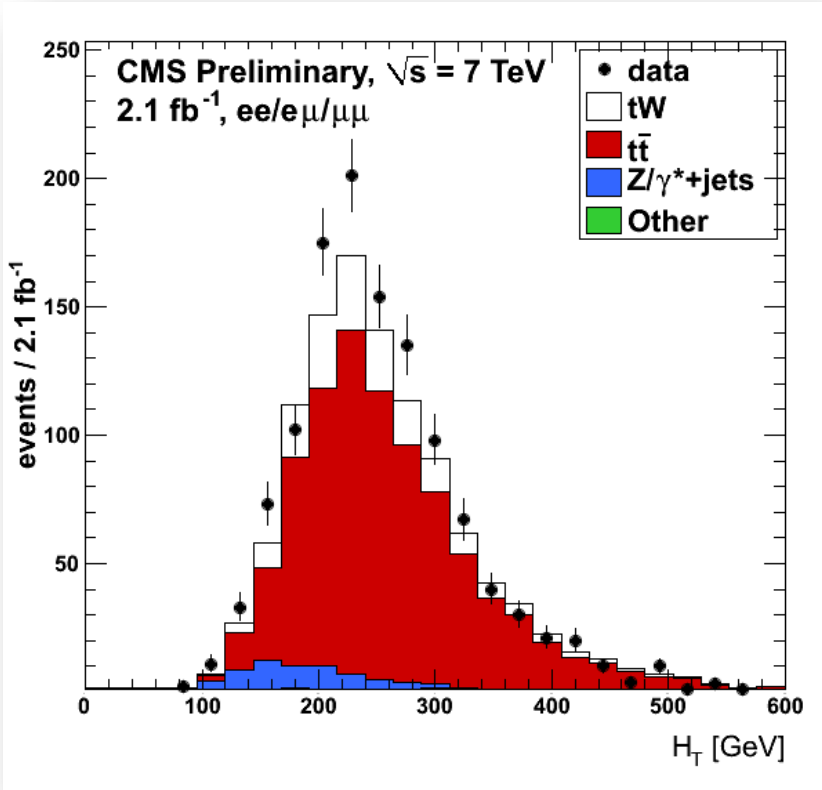
data and simulation after b-tagging in the ee, eμ and μμ final states separately and all together

Final MC/data distributions: H_T









data and simulation after b-tagging in the ee, eμ and μμ final states separately and all together

DY Background estimation

- WW, WZ, ZZ, W+jets, QCD and other single top processes contribute very little to the final number of events, and their contributions are taken from Monte Carlo

- In the signal region, **Drell-Yan is a main background** (ee/μμ), and it is estimated in a **data-driven** way (events in/out the Z window):

$$N^{estimated}_{ll,out} = \frac{N^{MC}_{ll,out}}{N^{MC}_{ll,in}} \left(N^{observed}_{ll,in} - \frac{1}{2} k \cdot N^{observed}_{e\mu,in} \right)$$

The number of observed eμ events is used to correct for the presence of **non-peaking backgrounds**, the factor 1/2 comes from the branching ratio and k corrects for the differences in acceptance and reconstruction between electrons and muons

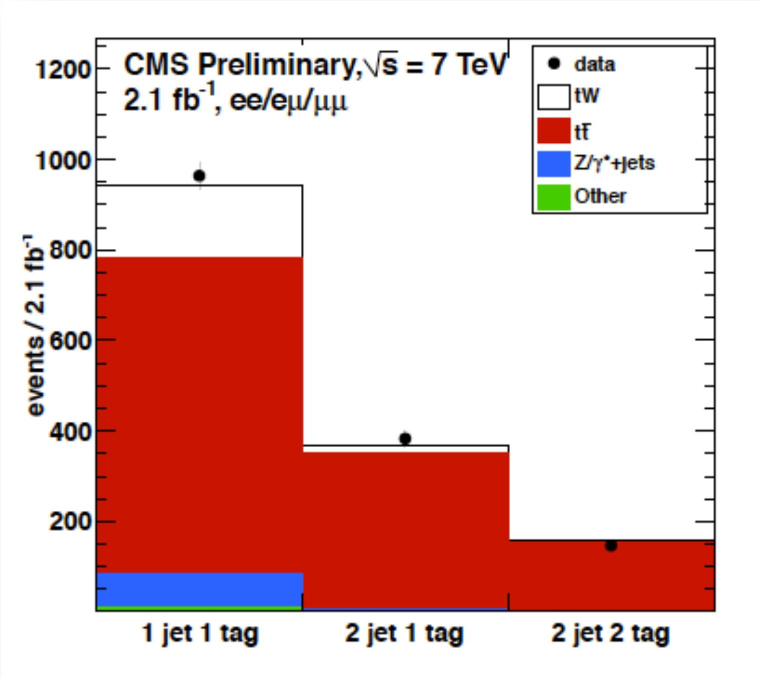
	ee channel	$\mu\mu$ channel
Data-driven estimate	20.7 ± 3.9	45.7 ± 6.1
Estimated from simulation	12±2	26±3

Uncertainty associated 50% (conservative)

- For the eμ final state (less important contribution), the simulation yield is also scaled to match the data by a factor 1.73, derived from the ee and μμ final states: 10.4±2.2 events (6.01.3 in simulation

TT control regions

- Two **tt enriched control regions:** 2jets1tag and 2jets 2tag are considered in the significance calculation to constrain tt contamination and b-tagging efficiency



Event yields in data and simulation in the signal region (1 jet, 1 b-tag) and the two tt-enriched control regions for the three final states together
Drell-Yan in ee/μμ comes from the data-driven estimation, and the **tt contribution is scaled to the outcome of the statistical fit**

Systematics (I)

- **Luminosity:** Luminosity measurement at CMS (4.5%)
- **Pile Up multiplicity:** Official Pile-up reweighting applied to the Monte Carlo, shifted by ±0.6 to estimate the systematic
- **Trigger Efficiency/Lepton reconstruction and identification efficiency:** Systematic taken directly from other dilepton analysis with similar selections (TOP-11-005/HIG-11-003)

- **MET modeling:** Scaling ±10% the un-clustered components of the MET

- **Jet multiplicity and jet energy scale (JES):** Shifting the jet energy correction (JEC) one standard deviation up/down, including the MET
- **Jet energy resolution (JER):** The JER is worse in data, everything is corrected by 10%, and to estimate the uncertainty, 0% for JER- and 20% for JER+

- **B-tagging:** Varying the b-tagging scale factor 0.95 by ±10%

Systematics (II)

- **Background Normalization:** Depends on the background, for DY, 50% comes from the data-driven estimation, for the rest backgrounds come from theory (uncertainty in the cross-section)

- **PDF uncertainties:** alternate weights for each event are produced using different pdf sets (CTEQ6.6 and MRST2006)

- **Factorization/Normalization Scale (Q2)**
- **ME/PS matching thresholds**
- **Initial and final state radiation (ISR/FSR)**

Dedicated MC samples with variations of the parameters up/down, doubled/halved... For tt and tW signal

- **DS/DR scheme:** Difference between the DR (default) and DS signal samples

- **Monte Carlo statistics:** Related to the size of the samples

Systematics (III)

Systematic uncertainty ($ee/e\mu/\mu\mu$) [%]	signal tW	$t\bar{t}$	$Z/\gamma^\star$	other
Luminosity	4.5	4.5	-	4.5
Pile-up multiplicity	0.48/0.55/0.73	$\star$	-	$\star$
Trigger Efficiency	1.5	1.5	-	1.5
Muon reconstruction and identification	- /1/1	- /1/1	-	- /1/1
Electron reconstruction and identification	2/2/ -	2/2/ -	-	2/2/ -
JES	$^{-2.5}_{+1.6}/^{-2.4}_{+0.1}/^{-0.6}_{+1.0}$	$^{-5.6}_{+4.4}/^{-6.0}_{+4.7}/^{-5.9}_{+2.3}$	-	$\star$
JER	1.1/0.5/0.4	3.1/3.9/4.4	-	$\star$
B-tagging	$^{-9.5}_{+10}/^{-9.8}_{+9.8}/^{-9.5}_{+10}$	$^{-8.5}_{+10}/^{-11}_{+10}/^{-9.1}_{+11}$	-	$\star$
Factorization/Normalization Scale (Q^2)	7.7/6/10	7.7/11/12	-	$\star$
ME/PS matching thresholds	-	5.7/0.7/2.3	-	$\star$
ISR/FSR	-	8.9/7.3/7.3	-	$\star$
DR/DS scheme	8.2/9.1/6.6	-	-	$\star$
E_T^{miss} modeling	2.3/0.9/0.9	$\star$	-	$\star$
PDF uncertainties	4.5/4.5/4.5	$\star$	-	$\star$
Background Normalization	-	15/15/15	50/ 50/ 50	$\star$
Simulation statistics	3.5/1.9/2.7	-	-	17/21/11

Rate systematic uncertainties for the three final states in the signal region, presented in percentage

"-" means it doesn't apply, and "★" for negligible contributions

Statistical Interpretation and Results

- The evaluation uses a statistical model of Poisson event counts in the three final states (ee, eμ, μμ)
- **Systematic uncertainties** are included in the model one by one as nuisance parameters (same treatment as in other multi-channel counting experiments such as Higgs searches)
- To estimate the tt contribution simultaneously, **events from tt enriched regions are included in the model**
- The significance is calculated with the background-only distribution of the likelihood ratio, and the 68% confidence level interval is evaluated using the profile likelihood method

Observed significance: **2.7** σ
Expected significance **1.8±0.9** σ

Cross section value and 68% C.L. interval: **22 +9-7 (stat + syst) pb**
(to compare with **15.6 ± 0.4 +1.0 -1.2 pb**)

Summary

- **First indications of the tW production in CMS** in the dilepton channel (three final states, ee, eμ and μμ) using **2.1fb^{-1}**
- Main background: **tt**
- Also important **Drell-Yan** (**data-driven** method)
- Smaller backgrounds (**WW/WZ/ZZ, single top, W+jets, QCD**) from simulation
- All the sources of systematic uncertainties addressed
- Signal region defined, and **significance calculated using the tt sidebands** (2jets1tag and 2jets2tag)

Observed (expected) **significance: 2.7σ** (1.8±0.9σ)

- Cross-section measured at the same time, value and 68% C.L. interval:

22 +9-7 (stat + syst) pb

- Coming up next:
  - Analyze the **full 2011 dataset**
  - Introduce Multivariate Methods to gain **sensitivity**